

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE MEDICINA VETERINÁRIA
CURSO DE ZOOTECNIA

ANNA CAROLINA FERNANDES

IDENTIFICAÇÃO DE CNVRs ASSOCIADAS AO TEOR DE COLESTEROL NA
CARNE DE BOVINOS DA RAÇA NELORE

UBERLÂNDIA – MG

2017

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE MEDICINA VETERINÁRIA
CURSO DE ZOOTECNIA

ANNA CAROLINA FERNANDES

Monografia apresentada à coordenação do curso graduação em Zootecnia da Universidade Federal de Uberlândia, como requisito parcial a obtenção do título de Zootecnista.

UBERLÂNDIA – MG

2017

ANNA CAROLINA FERNANDES

**IDENTIFICAÇÃO DE CNVRs ASSOCIADAS AO TEOR DE COLESTEROL NA
CARNE DE BOVINOS DA RAÇA NELORE**

Monografia apresentada como requisito parcial
para a aprovação na disciplina Trabalho de
Conclusão de Curso II do curso de graduação
em Zootecnia da Universidade Federal de
Uberlândia.

Banca Examinadora:

Dra. Fernanda Marcondes de Rezende – FAMEV/UFU
Presidente

Dra. Adriana Santana do Carmo – DPA/UFG
Membro

Dr. Robson Carlos Antunes – FAMEV/UFU
Membro

Uberlândia–MG

2017

AGRADECIMENTOS

Agradeço primeiramente a Deus por minha vida, família e amigos. Também à Universidade Federal de Uberlândia por ter disponibilizado todas as ferramentas que me permitiram chegar ao final desse ciclo de forma satisfatória.

Dedico esse trabalho aos meus mestres em geral, que me guiaram com sabedoria nesses anos de preparação, por sua grande contribuição nos caminhos por mim escolhidos e por despertarem em mim a boa ambição para com a Zootecnia, cujas referências e ensinamentos guardarei sempre comigo e com elevada consideração.

Uma palavra especial para minha querida orientadora Fernanda Marcondes de Rezende, que tanta ajuda forneceu para que este trabalho fosse concluído com o êxito esperado, pela oportunidade, confiança, paciência e empenho.

Da mesma forma não poderia deixar de agradecer minha família e em especial minha mãe e meu pai, presentes em todos os momentos desse percurso e que não me deixaram faltar palavras de incentivo e imensa compreensão em momentos de incerteza e dúvidas. Sem esse apoio seguramente não chegaria nesse momento.

Faço uma especial menção às professoras Adriana Santana do Carmo e Tatiane Cristina Seleguim Chud que também foram partes fundamentais para a realização deste trabalho, por todo o suporte e apoio.

Agradeço à banca examinadora pela contribuição à minha formação acadêmica.

Sou grata ao CNPQ e à FAPEMIG pelo apoio financeiro para a realização do trabalho.

Por fim, agradeço a todos que direta ou indiretamente fizeram parte desta etapa.

RESUMO

No Brasil, os bovinos da raça Nelore e os anelorados representam mais de 80 % dos animais que compõem o rebanho de corte nacional. Isso ocorre devido a características desse grupo de animais, como excelente adaptação ao ambiente tropical, resistência a endo e ectoparasitas e eficiência na conversão alimentar. Além disso, sua carne é caracterizada pelo baixo teor de gordura intramuscular (marmoreio) e uma distribuição homogênea da cobertura de gordura, característica valorizada no mercado. Porém, a ingestão de altos teores de gordura é prejudicial à saúde humana. O excesso de colesterol, por exemplo, é um fator associado a doenças cardiovasculares e síndromes metabólicas. Assim, objetiva-se a identificação de regiões de *copy number variations* (CNVRs) associadas ao teor de colesterol na carne bovina, o que possibilitará uma seleção melhor e mais eficiente de animais que produzam carne com melhores composições lipídicas. Para isso, biotecnologias desenvolvidas a partir do sequenciamento do DNA, como os *chips* de polimorfismos de nucleotídeo único (SNPs) foram usados para a identificação das CNVs e CNVRs. Os dados fenotípicos analisados foram mensurados em 615 animais, sendo a genotipagem realizada em apenas 407 animais pela plataforma Infinium com o ensaio Illumina Bovine beadchip HD® composto por 777.962 SNPs. A determinação dos genótipos foi realizada pelo software GenomeStudio v2011.1. Posteriormente, a identificação das CNVs foi efetuada por meio do programa PennCNV e a associação das CNVRs com o teor de colesterol ocorreu por regressão linear utilizando o programa CNVRuler. Quatro CNVRs foram detectadas e associadas de maneira favorável à deposição de colesterol na carne de bovinos da raça Nelore, ou seja, promovendo redução desse. Adicionalmente, foram identificados genes previamente descritos no genoma bovino nessas regiões CNV, cujo efeito está associado ao metabolismo lipídico em humanos.

Palavras-chave: Composição lipídica. Qualidade de carne. SNP *chip*. Variantes estruturais. Variação genética.

ABSTRACT

In Brazil, Nellore breed animals and their crosses represent over 80 % of all the animals that compose the beef herd. This is due to their excellent adaptation to tropical environment, resistance to endo and ectoparasites and efficiency in food conversion. In addition, its meat is characterized by low content of intramuscular fat (marbling) and they present a homogenous distribution of fat covering the carcass, which are traits valued on beef market. However, the ingestion of high levels of fat is harmful to human health. The excess of cholesterol, for example, is a factor associated to cardiovascular diseases and metabolic syndromes. Thus, our objective was the identification of copy number variation regions (CNVRs) associated with the cholesterol content in beef, which can be helpful for a more efficient selection of animals that produce meat with better lipid composition. For that, biotechnologies developed from the DNA sequencing, such as single nucleotide polymorphisms (SNPs) chips were used to identify CNVs and CNVRs. The phenotypic data were measured in 615 animals, being the samples genotyping performed in only 407 animals by Infinium platform with Illumina Bovine beadchip HD[®] composed of 777,962 SNPs. The genotypes were defined by GenomeStudio v2011.1. software. The CNVs identification was performed by the PennCNV program and the association between CNVRs with cholesterol content was done by linear regression using the CNVRuler program. Four CNVRs were detected with favorable association with the cholesterol deposition in Nellore beef, that means, promoting a reduction of it. In addition, genes previously described in the bovine genome were identified in these CNV regions and they were associated with lipid metabolism in humans.

Keywords: Lipid composition. Meat quality. SNP chip. Structural variants. Genetic variation.

SUMÁRIO

1 INTRODUÇÃO	07
2 REVISÃO DE LITERATURA	08
3 MATERIAL E MÉTODOS	11
3.1 Dados fenotípicos	11
3.2 Estimação dos efeitos genéticos e não genéticos	11
3.3 Extração de DNA e genotipagem	12
3.4 Identificação de CNVs e CNVRs	13
3.5 Correção para conteúdo de GC	15
4 RESULTADOS E DISCUSSÃO.....	16
5 CONCLUSÃO	21
REFERÊNCIAS	22
ANEXO I	25

1. INTRODUÇÃO

O estudo e o sequenciamento do genoma humano permitiram a descoberta de variantes estruturais que modificam a expressão dos genes. Os polimorfismos de nucleotídeo único (*single nucleotide polymorphisms* – SNPs) ainda são consideradas as principais alterações na estrutura da sequência de DNA, responsáveis por variações no fenótipo dos indivíduos (INTERNATIONAL HAPMAP CONSORTIUM, 2005). Entretanto, há alguns anos, descobriram-se outras variantes estruturais, as *copy number variations* (CNVs). Acredita-se que as CNVs possam influenciar a expressão dos genes, estando também relacionadas a variações fenotípicas (HASTINGS et al., 2009). Desse modo, aumentou o interesse no estudo do genoma de animais de produção pela possibilidade da associação dessas variantes com a expressão de características produtivas.

Ainda não se têm muitas informações acerca da contribuição efetiva de CNVs para o desempenho de bovinos. Com isso, a identificação dessas possibilitará a compreensão da arquitetura genética das características produtivas, permitindo a utilização destas mutações nos programas de melhoramento genético obtendo uma seleção mais eficiente de animais com melhores desempenhos. Sendo assim, é crucial realizar estudos de associação genômica ampla para identificar essas variantes estruturais e suas influências sobre características de produção e de qualidade do produto final.

Nesse contexto, objetivou-se, com o presente trabalho, identificar CNVs e CNVRs associadas ao teor de colesterol na carne de bovinos da raça Nelore.

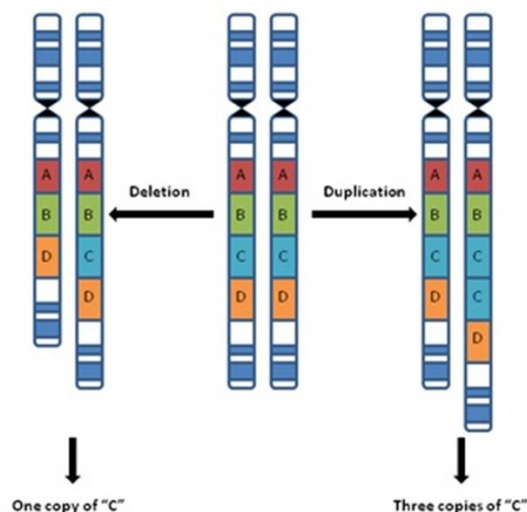
2. REVISÃO DE LITERATURA

As formas de variações genéticas mais frequentes no genoma são os polimorfismos do tipo SNP, os quais são resultado de processo natural ou induzido por agente externo, como agentes químicos ou radiação ionizante (BAE et al., 2010). Os SNPs são variações na sequência de DNA que ocorrem quando somente um nucleotídeo da sequência é diferente de um genoma de referência em pelo menos 1 % da população (INTERNATIONAL HAPMAP CONSORTIUM, 2005).

Acreditava-se que os SNPs fossem os principais responsáveis pela existência de variação fenotípica, tendo sido objeto de diversos estudos (INTERNATIONAL HAPMAP CONSORTIUM, 2005). Contudo, descobriu-se a existência de outra variante estrutural que tem sido foco dos estudos para a compreensão do controle genético da expressão das características: as CNVs (*copy number variations*, HASTINGS et al., 2009). Quando comparadas aos SNPs, considera-se que as CNVs têm maior efeito na expressão dos genes, além de apresentarem taxa de mutação mais elevada (ZHANG et al., 2014).

As CNVs (Figura 1) são alterações que ocorrem em fragmentos de DNA, podendo estas serem representadas por deleção, inserção, inversão ou translocação de genes (MCCARROLL et al., 2007). Pressupõe-se que as CNVs possam influenciar a expressão dos genes por estarem localizadas dentro ou próximo deles, podendo inclusive estar relacionadas a variações fenotípicas entre os indivíduos (MCCARROLL et al., 2007).

Figura 1. Representação dos tipos de CNVs.



Fonte: <http://readingroom.mindspec.org>

Pouco se sabe sobre a extensão da contribuição de CNVs para a variação genética em bovinos. Entretanto, CNVs já descritas em humanos, quando localizadas em regiões conservadas do genoma bovino, podem auxiliar nesses estudos (BICKHART et al., 2012). As CNVs nos bovinos estão distribuídos de forma aleatória e a probabilidade de cada cromossomo ser susceptível e/ou apresentar regiões que contêm CNVs varia de 1,32 – 8,80 % (FADISTA et al., 2010).

Com a disponibilidade de genomas de referência para bovinos (THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM, 2009; ZIMIN et al. 2009), novas possibilidades de pesquisa surgiram, entre elas o uso de ferramentas para detecção de SNPs que ajudam a identificar as CNVs; as quais podem ser utilizadas para acelerar o melhoramento genético de características complexas, com a possível associação entre as CNVs e a expressão dessas características (VAN TASSELL et al., 2008; MATUKUMALLI et al, 2009).

No Brasil, bovinos da raça Nelore representam quase a maioria dos animais que compõem o rebanho de corte (IBGE, 2014). Estudos comparativos entre raça Nelore e raças taurinas, como Angus e Hereford, têm demonstrado que animais Nelore apresentam, em sua maioria, menor número de CNVs quando comparados aos taurinos (PORTO NETO et al., 2011). Por outro lado, nota-se a presença duplicada de genes relacionados à melhor adaptação a regiões de clima quente, melhor resistência a patógenos e à infestação de parasitas em zebuínos (PORTO NETO et al., 2011), o que pode explicar a predominância destes animais no Brasil.

O mercado consumidor de carne bovina tem preferência por carne com pouca gordura de cobertura, visto a tendência crescente de grande parte da sociedade a buscar alimentos mais saudáveis, porém com gordura intramuscular, trazendo palatabilidade (PROENÇA, 2010). Não obstante e, levando isto em consideração, a ingestão controlada de colesterol, que é uma gordura, é essencial para a saúde, já que este nutriente atua em diversos processos como na formação das membranas celulares, na síntese de hormônios sexuais, tais como a testosterona, estrogênio e cortisol; na produção da bile, na formação da mielina, na metabolização de vitaminas lipossolúveis (A, D, E e K), entre outras funções (PORTAL BRASIL, 2012). O colesterol tem produção endógena, principalmente no fígado, porém faz-se necessária sua ingestão exógena, por meio dos alimentos, visto que a primeira não é suficiente para a realização de todas suas funções (SCHIAVO; LUNARDELLI; OLIVEIRA, 2003).

É importante ressaltar que altos índices de colesterol são muito prejudiciais à saúde. O colesterol é uma substância gordurosa que necessita de transportadores (lipoproteínas) para

sair da corrente sanguínea e atingir os tecidos periféricos (SILVA; COUTO, 2006). As principais lipoproteínas são o HDL (*High-Density Lipoprotein*), LDL (*Low-Density Lipoprotein*) e VLDL (*Very Low-Density Lipoprotein*). Estas duas últimas transportam colesterol e triglicerídeos do sangue para os tecidos. Já a lipoproteína HDL retira colesterol dos tecidos e devolve para o fígado, onde é degradado e, posteriormente, excretado pelos intestinos. Sendo assim, percebe-se que níveis aumentados de colesterol LDL são ruins para a saúde, assim como uma baixa concentração de colesterol HDL (SCHIAVO; LUNARDELLI; OLIVEIRA, 2003).

Além disso, o excesso de colesterol é um fator associado a inúmeras doenças, como deposição de gordura na parede dos vasos sanguíneos; doenças cardiovasculares e as síndromes metabólicas. Quase metade dos brasileiros apresenta índices elevados de colesterol (PORTAL BRASIL, 2012), esse fator, associado a grande incidência de outras doenças também relacionadas à alimentação, faz com que a população mude seus hábitos alimentares, e passe a adotar hábitos diferenciados e mais saudáveis, visando o aumento da expectativa de vida. Assim, com essa nova visão do mercado consumidor, fatores como a presença de elevado teor de colesterol influenciam muito na escolha dos alimentos.

Rezende (2008) observou que existe variabilidade nos teores de lipídios e de colesterol entre animais da mesma raça e sob mesmo regime alimentar, indicando portanto a existência de efeito genético sobre essas características e, assim a possibilidade de realização de melhoramento genético. Os programas de seleção de bovinos de corte priorizaram por muitos anos as características de crescimento e reprodutivas, e não deram muita ênfase nas características de qualidade da carne devido principalmente ao custo e a dificuldade de obtenção dos dados.

Espera-se que o desenvolvimento de novas tecnologias de sequenciamento genômico e de programas de análises complementares proporcionem melhores abordagens para identificar sistematicamente CNVs em todo o genoma (HOU et al., 2011). Também, espera-se que dessa forma a relação entre essas variantes e as características relativas à qualidade da carne em bovinos possa ser de fato assegurada e, futuramente incorporada nos programas de melhoramento genético, visando à obtenção de um produto final mais saudável.

O presente trabalho foi proposto com o objetivo de identificação das CNVs e a associação das CNVRs com a deposição de colesterol na carne de bovinos da raça Nelore.

3. MATERIAL E MÉTODOS

3.1. Dados fenotípicos

Os dados fenotípicos analisados foram mensurados em 615 animais da raça Nelore pertencentes ao programa de melhoramento genético da Agro-Pecuária CFM Ltda., cuja avaliação genética é realizada pelo Grupo de Melhoramento Animal e Biotecnologia da Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo (GMAB/FZEA/USP), localizado no município de Pirassununga, SP.

Esses animais foram criados em sistema extensivo até aproximadamente 20 meses de idade e, posteriormente, foram confinados por períodos de 90 a 150 dias, até atingirem aproximadamente 550 kg de peso vivo e serem destinados ao abate em abatedouro comercial, certificado para exportação para a Comunidade Comum Européia. Uma amostra do corte comercial denominado contrafilé, que corresponde ao músculo *Longissimus dorsi*, foi retirada de cada animal e identificada individualmente para a realização das análises de teor de colesterol. No momento do abate, foi avaliada a espessura de gordura subcutânea no corte realizado na altura da 12ª costela utilizando uma régua graduada em centímetros.

Os procedimentos para a extração do colesterol foram realizados no Laboratório de Bromatologia pertencente à FZEA/USP e a quantificação do teor de colesterol, por espectrofotometria, foi realizada no Laboratório de Morfofisiologia Molecular e Desenvolvimento, pertencentes à FZEA/USP. A determinação do teor de colesterol nas amostras de carne foi realizada de acordo com o método enzimático descrito por Saldanha, Mazalli e Bragagnolo (2004), o qual se baseia na degradação do colesterol pela enzima colesterol-oxidase produzindo peróxido de hidrogênio, que por meio de reação secundária produz cor. A intensidade da cor produzida é diretamente proporcional à quantidade de colesterol contida na amostra.

A estatística descritiva dos dados foi realizada por meio do procedimento PROC MEANS do pacote estatístico SAS® (SAS, 2011).

3.2. Estimação dos efeitos genéticos e não genéticos

As estimativas iniciais dos componentes de variância requeridos pelo MTDFREML (BOLDMAN et al., 1995) foram calculadas pelo procedimento PROC MIXED do pacote estatístico SAS® (SAS, 2011), considerando, no modelo de análise, os lotes de abate e de

análise do teor de colesterol como efeitos fixos classificatórios e a espessura de gordura subcutânea e a idade ao abate como efeitos fixos contínuos, bem como o efeito aleatório de touro.

A estimação dos efeitos genéticos e não genéticos sobre o teor de colesterol foi realizada pelo programa MTDFREML (BOLDMAN et al., 1995), sob modelo animal, considerando os mesmos efeitos fixos descritos anteriormente e a matriz de parentesco composta por 4.065 animais. O critério de convergência adotado para a conclusão das análises foi quando a variância do simplex atingiu 10^{-12} , após ter convergido a 10^{-6} e 10^{-9} em rodadas anteriores.

O fenótipo ajustado, calculado com base nas estimativas dos efeitos não genéticos obtidas no MTDFREML, foi utilizado nas análises de associação com as CNVRs no programa CNVRuler (KIM et al., 2012), o qual não aceita valores negativos, impossibilitando a utilização dos valores genéticos estimados como fenótipo para a realização das análises.

3.3. Extração de DNA e genotipagem

A extração do DNA genômico foi realizada a partir de amostras de carne utilizando o kit comercial NucleoSpin[®] Tissue (Macherey-Nagel) de acordo com o protocolo preconizado pelo fabricante, disponível em detalhes no endereço eletrônico <http://www.mn-net.com/>. A concentração (em ng/mL) e o grau de pureza (relação entre as densidades óticas a 260 e 280nm) do DNA foram mensurados utilizando-se espectrofotômetro (Nanodrop - Thermo Fisher Scientific, Waltham, Massachusetts, EUA).

Dentre os animais avaliados para o teor de colesterol, apenas 407 foram genotipados pela plataforma Infinium com o ensaio Illumina Bovine beadchip HD[®] composto por 777.962 SNPs. Para tanto, seguiu-se o protocolo fornecido pelo fabricante e a determinação dos genótipos foi realizada pelo software GenomeStudio v2011.1, no qual é gerado o arquivo FinalReport (Figura 2) usado para a identificação das CNVs.

Figura 2. Modelo do arquivo FinalReport gerado pelo software GenomeStudio.

```

GGST Version 1.9.4
Processing Date 3/12/2014 1:51 PM
Content BovineHL_B.bpm
Num SNPs 777962
Total SNPs 777962
Num Samples 407
Total Samples 407
File 1 of 407
[Data]
SNP Name Sample ID Allele1 - Forward Allele2 - Forward Allele1 - Top Allele2 - Top Allele1 - AB Allele2 - AB Log R Ratio B Allele Freq GC Score
ARS-BFGL-BAC-10172 185A G G G G B B -0.2340 0.9969 0.9374
ARS-BFGL-BAC-1020 185A A G A G A B 0.0780 0.5145 0.9568
ARS-BFGL-BAC-10245 185A C C G G B B -0.0916 0.9645 0.7996
ARS-BFGL-BAC-10345 185A A A A A A A 0.0105 0.0000 0.9604
ARS-BFGL-BAC-10365 185A G G C C B B 0.1585 0.9908 0.5296
ARS-BFGL-BAC-10591 185A A A A A A A -0.3826 0.0000 0.8934
ARS-BFGL-BAC-10793 185A C C G G B B -0.3173 1.0000 0.9549
ARS-BFGL-BAC-10867 185A G G C C A A -0.3539 0.0172 0.9592
ARS-BFGL-BAC-10951 185A T T A A A A 0.8282 0.0000 0.5453
ARS-BFGL-BAC-10952 185A A A A A A A -0.0343 0.0122 0.9274
ARS-BFGL-BAC-10960 185A G G G G B B 0.0277 0.9816 0.9529
ARS-BFGL-BAC-10972 185A C C G G B B 0.1225 0.9997 0.9195
ARS-BFGL-BAC-10975 185A G G G G B B 0.3007 0.9935 0.8055
ARS-BFGL-BAC-11000 185A T T A A A A -0.2630 0.0000 0.9575
ARS-BFGL-BAC-11003 185A C C G G B B 0.0058 0.9991 0.9177
ARS-BFGL-BAC-11007 185A T T A A A A 0.2479 0.0002 0.8796

```

As plataformas de genotipagem utilizam um padrão de intensidade de sinal esperado para cada posição que é dada pelo log R. Baseado nisso, primeiro é feita a identificação do genótipo AA (2 vermelhos), AB (vermelho e verde) ou BB (2 verdes). Existe uma variação aceita na intensidade de cada posição, se a intensidade do sinal está dentro desse limite, então diz-se que houve *call* para o SNP. Caso a intensidade de sinal esteja fora do limite diz-se *no call* para o SNP, portanto não é possível genotipar. Uma vez definido o genótipo AA, AB ou BB, utiliza-se o banco de dados da plataforma para identificar qual a base correspondente à posição (A, C, T ou G) e, por fim, determinar os genótipos (ex. CC, CT ou TT).

3.4. Identificação de CNVs e CNVRs (*copy number variations regions*)

A identificação das CNVs foi realizada por meio do programa PennCNV desenvolvido por Wang et al. (2007), que é uma ferramenta gratuita, disponível para download no endereço eletrônico <https://github.com/WGLab/PennCNV/releases> e que trabalha com dados de intensidade de sinal dos ensaios Illumina e Affymetrix. O software utiliza as métricas Log R Ratio (LRR) e a frequência do alelo B (BAF, do inglês, *B allele frequency*), provenientes do processo de genotipagem (Figura 2). O LRR indica a intensidade de sinal total normalizada de alelos e o BAF representa os desequilíbrios alélicos, ou seja, a proporção normalizada do alelo B no genótipo (POPOVA et al., 2009).

Uma região cromossômica normal tem três possíveis agrupamentos de genótipos BAF, representados como AA, AB e BB, e com valores LRR centrados em torno de zero. O aumento do número de cópias para uma região CNV pode ser detectado com base no aumento

do número de picos na distribuição de BAF, bem como o aumento dos valores LRR. Assim, para indivíduos que apresentam menos ou mais cópias, o valor LRR é esperado ser negativo e positivo, respectivamente (WANG et al., 2007).

O processo de detecção de CNVs pelo programa PennCNV é realizado em várias etapas, por meio de diferentes rotinas computacionais (*scripts*). Uma breve descrição dos *scripts* utilizados é apresentada a seguir:

- `split_illumina_report`: gera arquivo individual de intensidade de sinal com informações de LRR e BAF para cada animal;
- `compile_pfb`: gera arquivo pfb (do inglês, *population frequency of allele B*), que contém nome do SNP, cromossomo, posição e frequência populacional do alelo B, que serão utilizados pelo programa `detect_cnv`. A frequência do alelo B é calculada para cada SNP na população e, depois, é usada como referência da população analisada para o controle de qualidade;
- `detect_cnv`: detecta as CNVs utilizando o HMM (do inglês, *Hidden Markov Model*) para a solução de modelos complexos por meio da abordagem probabilística (EDDY, 2004). Quando se tem CN (*copy number*) = 2 indica normal/estado neutro (duas cópias do gene porque é diploide); CN = 0 ou 1 indica deleção e CN = 3 ou 4 indica 1 ou 2 duplicações, porém não se sabe em qual cromossomo. Os cromossomos sexuais X e Y não foram considerados na análise.
- `filter_cnv`: faz controle de qualidade das amostras a partir de critérios padrão (*default*), sendo o argumento `-qclrrsd` (controle de qualidade do desvio padrão LRR) igual a 0,3, o argumento `-qcbafdrift` (controle de qualidade da deriva de BAF) igual a 0,01 e o argumento `-qcwf` (controle de qualidade do fator de ondulação) igual a 0,05. O `-qclrrsd` representa o limite máximo de desvio padrão LRR, o `-qcbafdrift` representa o limite máximo de deriva de BAF e o `-qcwf` representa o limite de intervalo de fator de ondulação para inclusão da amostra na análise. O fator de ondulação mede a flutuação da intensidade de sinal, que é altamente correlacionada ao conteúdo de GC, nos ensaios de detecção de SNPs prejudicando sua correta inferência (DISKIN et al., 2008). O conteúdo de GC representa a porcentagem de bases nitrogenadas guanina e citosina em moléculas de DNA ou RNA. Sabe-se que regiões com um baixo conteúdo de GC correspondem aos picos da onda enquanto as regiões com elevado teor de GC correspondem às depressões (MARIONI et al., 2007). Além dos critérios padrão utilizados, foram considerados para o controle de qualidade outros três parâmetros: –

numsnps 10, -qcnumCNV 150 e -maxlength 5Mb. Em seguida, utilizou-se o pacote estatístico R para checar as estatísticas utilizadas no controle de qualidade de acordo com os parâmetros estimados nos dados (ANEXO I).

As CNVRs foram identificadas por meio do programa CNVRuler (KIM et al., 2012), quando observado ao menos um par de base de sobreposição entre as CNVs. As CNVRs foram classificadas com base nos eventos (ganho ou perda de cópias) determinados pelo PennCNV (OLIVEIRA-JÚNIOR et al., 2014), sendo que as CNVRs denominadas mistas, representam as regiões genômicas em que ambos os eventos podem ocorrer (SILVA, 2016). Simultaneamente, o CNVRuler realiza a associação das CNVRs com o teor de colesterol por análise de regressão linear, sendo consideradas como CNVRs apenas aquelas que apresentaram frequência acima de 5 % na população, pois abaixo disso são tidas como raras (WANG et al., 2015), podendo representar erros no processo de identificação.

Por fim, as CNVRs com efeito significativo ($p < 0,15$) no teor de colesterol foram investigadas no Ensembl, disponível no endereço eletrônico <http://www.ensembl.org/index.html>, para busca de genes já descritos, utilizando como referência o genoma bovino (Cow UMD 3.1).

3.5. Correção para conteúdo de GC

A ondulação da intensidade de sinal causada pela presença de GC e responsável por alterar o padrão de intensidade de sinal esperado para cada genótipo (SNP) é mais comum de acontecer em humanos, possivelmente pela localização das *probes* escolhidas para a genotipagem. Para efetuar a correção para o conteúdo de GC, foi utilizado o conteúdo de GC do genoma de referência bovino (Cow UMD 3.1). Esse ajuste foi realizado na fase do programa detect_cnv do PennCNV e tem como principal objetivo reduzir o número de erros de identificação de mutações devido ao efeito de ondulação. Os demais procedimentos e controles para detecção de CNVs e CNVRs descritos anteriormente foram mantidos.

Embora, outros estudos não tenham identificado alta influência do ajuste para conteúdo de GC em bovinos na detecção de CNVs (comunicação verbal CARMO e CHUD, 2017), no presente estudo, o ajuste foi realizado para avaliar seu impacto na identificação de CNVRs associadas ao teor de colesterol.

4. RESULTADOS E DISCUSSÃO

As estatísticas descritivas para a característica teor de colesterol na carne de machos jovens inteiros da raça Nelore estão apresentadas na Tabela 1.

Tabela 1. Estatísticas descritivas para quantidade de colesterol em 100 g de contrafilé de bovinos da raça Nelore.

Característica	N	MIN	MAX	MED	DP	CV (%)
COLESTEROL, mg/100g	615	28,76	83,95	56,42	8,26	14,65

N = número de animais; MIN = valor mínimo observado; MAX = valor máximo observado; MED = média; DP = desvio padrão; CV = coeficiente de variação.

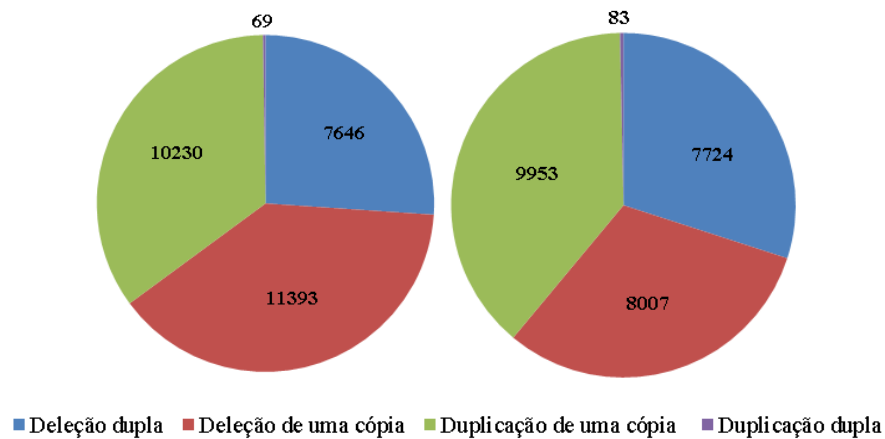
De acordo com a Tabela Brasileira de Composição de Alimentos (2011), a quantidade de colesterol por 100 g de uma carne bovina contrafilé cru sem gordura de cobertura é de 59 mg, o que indica que os dados estão em conformidade com a literatura, já que a média obtida foi de 56 mg, com desvio padrão de 8 mg. Os valores de mínimo e máximo apresentaram uma variação considerável já que existem animais com 28 mg de colesterol, ao passo que outros apresentam 84 mg em 100 g de contrafilé.

Em relação às estimativas dos componentes de variância e do coeficiente de herdabilidade para o teor de colesterol, foram obtidos os seguintes resultados: variância aditiva igual a 0,37 mg²/100 g²; variância residual igual a 36,89 mg²/100 g²; variância fenotípica igual a 37,26 mg²/100 g² e coeficiente de herdabilidade igual a 0,01. A estimativa do componente de variância aditiva demonstra a existência de variabilidade genética aditiva, a qual é indicativa da possibilidade de seleção para essa característica. Entretanto, além do coeficiente de herdabilidade estimado ser extremamente baixo, essa característica é de mensuração tardia, cara e depende do abate do animal para ser avaliada, o que justifica o uso da informação genômica como ferramenta de otimização para seleção e melhora genética para essa característica, ao invés do melhoramento genético clássico baseado exclusivamente no fenótipo e na genealogia (MEUWISSEN; HAYES; GODDARD, 2001).

No processo de identificação de CNVs duas abordagens foram adotadas: sem e com o ajuste para o conteúdo de GC. Na primeira, foram encontradas 29.338 potenciais CNVs (sendo 9.233 CNVs únicas), representando 7.647 deleções duplas, 11.393 deleções de uma cópia, 10.230 duplicações de uma cópia e 69 duplicações duplas, nas 407 amostras avaliadas e, o número de SNPs por CNV variou de 3 a 2.113 (Figura 3a). Ao ser realizado o controle para conteúdo de GC, 25.767 potenciais CNVs (sendo 7.053 CNVs únicas), representando

7.724 deleções duplas, 8.007 deleções de uma cópia, 9.953 duplicações de uma cópia e 83 duplicações duplas foram encontradas nas 407 amostras avaliadas e, o número de SNPs por CNV variou de 3 a 751 (Figura 3b).

Figura 3. Frequência das CNVs identificadas por tipo de alteração no DNA.



Na Figura 4, está representada a distribuição cromossômica de todas as CNVs identificadas e, na Figura 5, tem-se a distribuição dos tamanhos dessas CNVs. As CNVs identificadas nas abordagens sem e com ajuste para o conteúdo de GC totalizaram 2.088.701.802 e 1.765.895.001 pares de bases (pb), representando 69,62 e 58,86 %, respectivamente, do genoma bovino, que é composto por aproximadamente 3 bilhões de pb.

Figura 4. Distribuição cromossômica das CNVs detectadas nos dados de genotipagem de SNP de alta densidade em bovinos da raça Nelore.

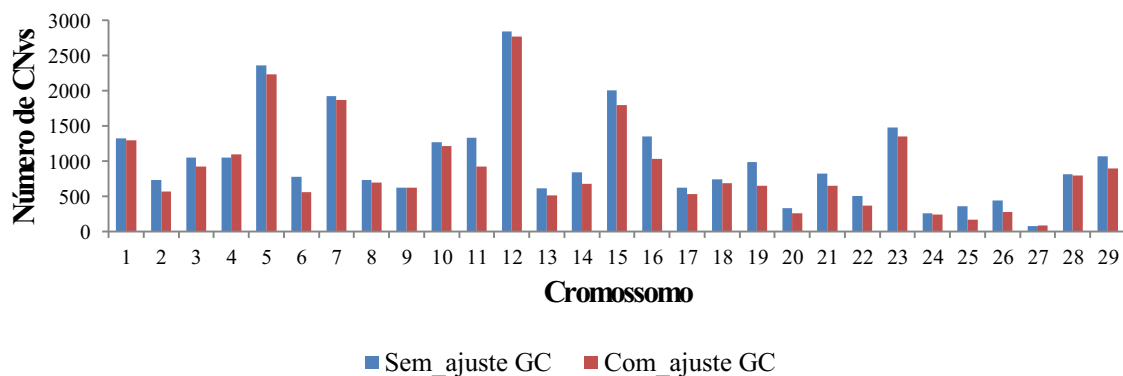
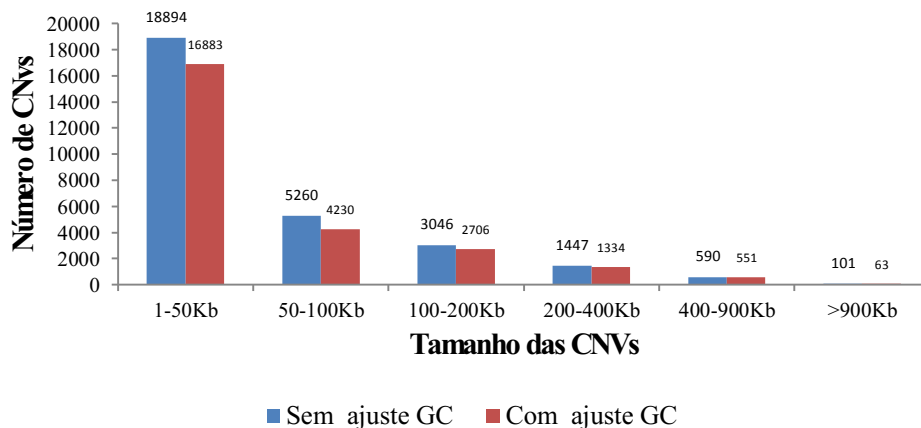


Figura 5. Distribuição do tamanho das CNVs detectadas nos dados de genotipagem de SNP de alta densidade em bovinos da raça Nelore.



As distribuições cromossômica (Figura 3) e do tamanho (Figura 4) das CNVs identificadas pelo processo de genotipagem com *chip* de SNP de alta densidade no presente trabalho diferiram das descritas por Silva et al. (2016). Essas variações podem ser explicadas pelas divergências nas populações analisadas. No referido estudo, os autores utilizaram para genotipagem amostras de sêmen comercialmente disponíveis, contrapondo ao atual que avaliou animais pertencentes a uma população submetida a intenso programa de seleção.

Após o controle de qualidade, restaram 8.214 CNVs (sendo 2.798 CNV únicas) em 352 amostras e 7.903 CNVs (sendo 2.503 CNV únicas) em 375 amostras nas abordagens sem e com ajuste para o conteúdo de GC, respectivamente. Ao serem condensadas as CNVs que apresentaram sobreposição de pelo menos um par de bases, 58 e 63 CNVRs, com tamanho variando de 30,8 Kpb – 2,9 Mpb, foram identificadas, das quais 10 e 12 CNVRs foram associadas à deposição de colesterol ($p < 0,15$) nas abordagens sem e com ajuste para conteúdo de GC, nessa ordem. As CNVRs identificadas representaram 0,80 e 0,86 % do genoma bovino (23.717.936 e 25.789.337 pb).

As CNVRs associadas ao teor de colesterol pelas duas abordagens consideradas (sem e com ajuste para conteúdo de GC) estão descritas na Tabela 2.

Tabela 2. CNVRs associadas ao teor de colesterol na carne de bovinos Nelore, identificadas sem e com ajuste para o conteúdo de GC no genoma bovino.

GC	Chr	Início (pb)	Fim (pb)	Freq (%)	p-valor	Estimativa (EP)	Tipo
Sem	1	105001726	105270633	10,51	0,13	-3,31±4,27	Perda
	11	5745177	5800039	12,50	0,01	5,26±3,94	Ganho
	11	103939132	104185394	16,76	0,10	2,93±3,51	Perda
	13	12543348	12822615	15,91	0,01	-4,88±3,56	Ganho
	13	71850682	71931812	11,65	0,08	3,61±4,08	Perda
	14	74984	1410699	29,26	0,12	-2,32±2,88	Ganho
	18	49074618	49109287	11,08	0,004	6,12±4,14	Perda
	18	13388207	13445868	9,38	0,05	4,56±4,49	Ganho
	19	1591038	1683625	14,49	0,03	-4,25±3,71	Perda
	19	51934105	52272577	12,50	0,11	3,26±3,97	Perda
Com	1	105001726	105270633	11,65	0,12	-3,21±4,03	Perda
	4	113657341	114075519	25,87	0,06	2,77±2,87	Ganho
	7	42945525	43353211	22,40	0,11	2,44±3,02	Perda
	10	40734273	40794738	8,53	0,01	-5,9±4,47	Perda
	11	5745177	5800039	14,93	0,03	3,85±3,51	Ganho
	13	12543348	12822615	15,20	0,01	-4,63±3,48	Ganho
	13	71850682	71931812	9,33	0,08	3,84±4,32	Perda
	14	74984	1410699	37,33	0,06	-2,46±2,59	Ganho
	15	43737	508576	14,67	0,05	3,61±3,54	Ganho
	18	13350936	13455012	12,00	0,03	4,22±3,86	Ganho
	18	57047421	57194210	9,07	0,04	4,61±4,36	Ganho
	19	1591038	1691733	15,47	0,09	-3,01±3,47	Perda

Chr: cromossomo bovino; pb: pares de bases; Freq: frequência das CNVRs; EP: erro-padrão da média.

Analisando a Tabela 2, nota-se que seis CNVRs, localizadas nos cromossomos 1, 11, 13, 14 e 19, foram identificadas associadas ao teor de colesterol, independente do ajuste para o conteúdo de GC. Sendo que, quatro delas, representadas pelas CNVRs presentes nos cromossomos 1, 13 (12543348-12822615 pb), 14 e 19 (1591038-1691733 pb) apresentaram estimativas favoráveis, ou seja, promovem redução na deposição de colesterol na carne de bovinos da raça Nelore.

A investigação das regiões CNV significativas, em ambas abordagens, permitiu a detecção de genes previamente descritos no genoma bovino, os quais estão apresentados na Tabela 3.

Tabela 3. Associação de CNVRs com genes previamente descritos.

Gene	Chr/Nome	Início (pb)	Fim (pb)	Gene	Tipo
ENSBTAG000000047958	GJ059509.1	3535	6470	DPP7	codifica proteína
ENSBTAG000000020495	GJ058424.1	10382	37633	SH3YL1	codifica proteína
ENSBTAG000000003540	GJ058437.1	1	37340		codifica proteína
ENSBTAG000000035144	GJ060118.1	1942	8662		codifica proteína
ENSBTAG000000038537	20	13234	15549		codifica proteína
ENSBTAG000000038364	20	56836	57927		codifica proteína
ENSBTAG000000000617	20	67609	68945		codifica proteína
ENSBTAG000000045490	20	107180	108069		pseudogene
ENSBTAG000000047333	20	132123	133868		codifica proteína
ENSBTAG000000011895	20	182976	204785	PANK3	codifica proteína

Os três genes identificados tiveram seu efeito associado ao metabolismo lipídico em estudos prévios com humanos. O gene DPP7, segundo Cohen et al. (1999), está envolvido com o transporte de lipídios por meio da atividade de um transportador transmembrana, o gene SH3YL1 modula uma ligação entre componentes dos lipídios (HASEGAWA et al., 2011) e o gene PANK3 parece participar de vias metabólicas lipídicas (WILFRED et al., 2007). Nesse sentido, estudos complementares deverão ser desenvolvidos para a melhor compreensão da função desses genes na regulação do colesterol na carne de bovinos da raça Nelore.

5. CONCLUSÃO

No presente trabalho, CNVs foram identificadas e CNVRs foram associadas ao teor de colesterol na carne de bovinos da raça Nelore. Quatro CNVRs, presentes nos cromossomos 1 (105001726-105270633), 13 (12543348-12822615 pb), 14 (74984-1410699) e 19 (1591038-1691733 pb), parecem reduzir a deposição do colesterol. Nas regiões CNVs foram identificados genes (DPP7, SH3YL1 e PANK3) previamente descritos no genoma bovino que estão relacionados ao metabolismo lipídico em humanos.

REFERÊNCIAS

- BAE J.S., CHEONG H.S., KIM L.H., NAMGUNG S., PARK T.J., CHUN J.Y., KIM J.Y., PASAJE C.F., LEE J.S., SHIN H.D. Identification of copy number variations and common deletion polymorphisms in cattle. **BMC Genomics** v.11, p.232, 2010.
- BICKHART D.M., HOU Y., SCHROEDER S.G. et al. Copy number variation of individual cattle genomes using next-generation sequencing. **Genome Res.** v.22, n.4, p.778–790, 2012.
- BOLDMAN, K.G.A.; KRIESE, L.D.; VAN VLECK, L.P. et al. A manual for use of MTDFREML. A set of programs to obtain estimates of variance and covariances. **ARS, USDA**, Washington, DC, 1995.
- THE BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM. The Genome Sequence of Taurine Cattle: A window to ruminant biology and evolution. **Science** v.324, p.522–528, 2009.
- COHEN D.E., GREEN R.M., WU M.K., BEIER D.R. Cloning, tissue-specific expression, gene structure and chromosomal localization of human phosphatidylcholine transfer protein. **Biochim Biophys Acta.** p. 265–70, 1999.
- DISKIN S. J., LI M., HOU C., YANG S., GLESSNER J., HAKONARSON H., BUCAN M., MARIS J. M., WANG K. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. **Nucleic Acids Res.**, v.36, 2008.
- EDDY S. R. What is a hidden Markov model? **Nature Biotechnology**, p 1315-6, 2004.
- FADISTA J., THOMSEN B., HOLM L.E., BENDIXEN C.: Copy number variation in the bovine genome. **BMC Genomics**, 2010.
- HASEGAWA J., TOKUDA E., TENNO T., TSUJITA K., SAWAI H., HIROAKI H., TAKENAWA T., ITOH T. SH3YL1 regulates dorsal ruffle formation by a novel phosphoinositide-binding domain. **J. Cell Biol.** v. 193, nº 5, p. 901–916, 2011.
- HASTINGS P.J., LUPSKI J.R., ROSENBERG S.M., IRA G. Mechanisms of change in gene copy number. **Nat Rev Genet** v.10, p.551–564, 2009.
- HOU Y., LIU G.E., BICKHART D.M., CARDONE M.F., WANG K., KIM E.S., MATUKUMALLI L.K., VENTURA M., SONG J., VANRADAN P.M., et al. Genomic characteristics of cattle copy number variations. **BMC Genomics**, 2011.
- IBGE. **PPM 2010: Rebanho bovino nacional cresce 2,1 % e chega a 209,5 milhões de cabeças.** 2014. Disponível em:<http://www.ibge.gov.br/home/presidencia/noticias/noticia_visualiza.php?id_noticia=2002&id_pagina=1>. Acesso em: 22 out. 2016.
- INTERNATIONAL HAPMAP CONSORTIUM: A haplotype map of the human genome. **Nature.** v.437, p.1299-1320, 2005.
- KIM, J. H., HU, H. J., YIM, S. H., et al. CNVRuler: a copy number variation-based case-control association analysis tool. **Bioinformatics** v.28, p.1790–1792, 2012.

MARIONI J.C., THORNE N.P., VALSESIA A., FITZGERALD T., REDON R., FIEGLER H., ANDREWS T.D., STRANGER B.E., LYNCH A.G., DERMITZAKIS E.T., CARTER N.P., TAVARÉ S., HURLES M. E. Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. **Genome Biol.**, v.8, 2007.

MATUKUMALLI L.K., LAWLEY C.T., SCHNABEL R.D., TAYLOR J.F., ALLAN M.F., HEATON M.P., et al: Development and characterization of a high density SNP genotyping assay for cattle. **PLoS ONE**. 2009

MCCARROLL S.A., ALTSHULER D.M.: Copy-number variation and association studies of human disease. **Nat Genet** v.39, p.37-42, 2007.

MEUWISSEN T.H.E., HAYES B. J., GODDARD M.E. Using the genomic relationship matrix to predict the accuracy of genomic selection. **J. Anim. Breed. Genet.** v.128, p.409 – 421, 2011.

OLIVEIRA JÚNIOR G. A., F. REY S. B., SANTANA M. H. A., FERRAZ J. B. S., et al. **Analysis of copy number variation regions in a Nellore population evaluated for feed efficiency.** Proceedings, 10th World Congress of Genetics Applied to Livestock Production, 2014.

POPOVA T., MANIÉ E., STOPPA-LYONNET D., RIGAILL G., BARILLOT E., STERN M.H. Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. **Genome Biol**, 2009.

PORTAL BRASIL. **Colesterol é uma das principais causas de problemas cardíacos no país.** 2012. Disponível em: <<http://www.brasil.gov.br/saude/2012/08/dia-nacional-do-controle-do-colesterol-e-comemorado-nesta-quarta-8>>. Acesso em: 21 maio de 2015.

PORTO NETO L.R., JONSSON N.N., D'OCCHIO M.J., BARENDSE W. Molecular genetic approaches for identifying the basis of variation in resistance to tick infestation in cattle. **Vet Parasitol** v.180, p.165–172, 2011.

PROENÇA R.P.C. Alimentação e globalização: algumas reflexões. **Cienc. Cult.** v.62 no.4 São Paulo, 2010

REZENDE, F. M. **Prospecção da influência de marcadores genéticos sobre características de crescimento, carcaça e qualidade de carne em bovinos da raça Nelore.** 2008. 171 f. Dissertação (Mestrado) – Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2008.

SALDANHA, T; MAZALLI, M.R.; BRAGAGNOLO, N. Avaliação comparativa entre dois métodos para determinação do colesterol em carnes e leite. **Ciência e Tecnologia de Alimentos**, v. 24, n. 1, p. 109-113, 2004.

SAS. Base SAS® 9.3 Procedures Guide: **Statistical Procedures**. Cary: SAS, 2011. 536p

SCHIAVO, M.; LUNARDELLI, A.; OLIVEIRA, J.R. Influência da dieta na concentração sérica de triglicerídeos. **Jornal Brasileiro de Patologia e Medicina Laboratorial**. Rio de Janeiro, v. 39, n. 4, p. 283-288, 2003.

SILVA E.L., COUTO, R. D. Estrutura, metabolismo e funções fisiológicas da lipoproteína de alta densidade. **Jornal Brasileiro de Patologia e Medicina Laboratorial**. Rio de Janeiro, Brasil, v. 42, no. 3, p. 169-178, 2006.

SILVA V.H., REGITANO L.C.A., GEISTLINGER L., PÉRTILLE F., GIACHETTO P.F., BRASSALOTI R.A., MOROSINI N.S., ZIMMER R., COUTINHO L.L. Genome-Wide Detection of CNVs and Their Association with Meat Tenderness in Nelore Cattle. **PLoS ONE**, 2016.

TABELA BRASILEIRA DE COMPOSIÇÃO DE ALIMENTOS - TACO. 4. ed. rev. e ampl. Campinas, SP: NEPA- UNICAMP, 161 p., 2011.

VAN TASSELL C.P., SMITH T.P., MATUKUMALLI L.K., TAYLOR J.F., SCHNABEL R.D., LAWLEY C.T., et al: SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. **Nat Methods**. v.5, p.247-252, 2008.

WANG, K., LI, M., HADLEY, D., et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. **Genome Res**. v.17, p.1665–1674, 2007.

WANG WB., WANG W., SUN W., CROWLEY J. J., SZATKIEWICZ J. P. Allele-specific copy-number discovery from whole-genome and whole-exome sequencing. **Nucleic Acids Res.**, v.43, 2015.

WILFRED B.R., WANG WX., NELSON P.T. Energizing miRNA research: A review of the role of miRNAs in lipid metabolism, with a prediction that miR-103/107 regulates human metabolic pathways. **Science Direct**, v. 91, no. 3, p. 209–217, 2007.

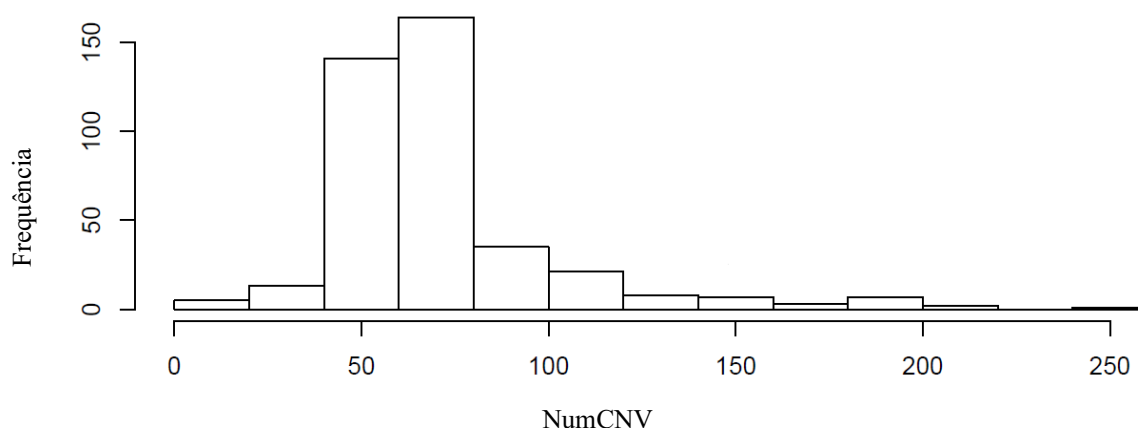
ZHANG F., JIA S.; YANG M.; XU Y.; LIU C., et al. Detection of copy number variations and their effects in Chinese bulls. **BMC Genomics** v.15, p.480, 2014.

ZIMIN A.V., DELCHER A.L., FLOREA L., KELLEY D.R., SCHATZ M.C., PUIU D., HANRAHAN F., PERTEA G., VAN TASSELL C.P., SONSTEGARD T.S., et al. A whole-genome assembly of the domestic cow, *Bos taurus*. **Genome Biol** v.1, p.42, 2009.

ANEXO I

O argumento `-numsnp` representa o número mínimo de SNPs presentes nas CNVs por amostra; tendo sido removidas as CNVs com menos de 10 SNPs por amostra com o objetivo de reduzir as taxas falso negativas obtidas de CNVs muito pequenas (<http://penncnv.openbioinformatics.org/en/latest/>). O argumento `-qnumcnv` (controle de qualidade para o número de CNVs) foi programado para `'qnumcnv 150'`. Assim, qualquer amostra com mais de 150 CNVs seria uma amostra de baixa qualidade e portanto eliminada da análise com o objetivo de não inflar o comprimento da CNVR devido à concatenação de CNVs falso positivas detectadas (<http://penncnv.openbioinformatics.org/en/latest/misc/faq/>), porém a distribuição observada no histograma gerado pelo software R foi considerada para decidir acerca de um bom limite a ser utilizado (Gráfico 1).

Gráfico 1. Histograma NumCNV



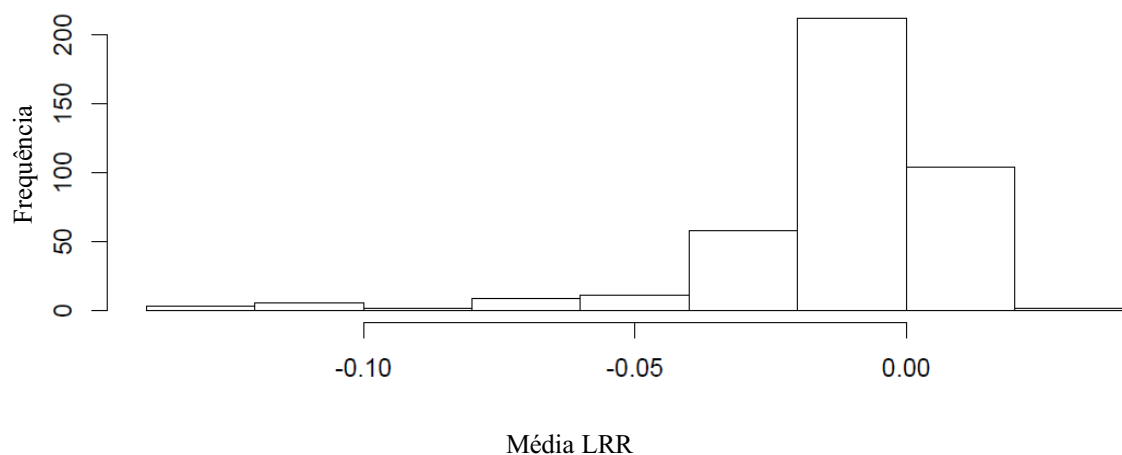
Dessa forma, considerando que valores acima do limite superior ($-Q3+1,5df$) devem ser removidos, por se tratar de *outliers*, pelo boxplot deveriam ser eliminadas as amostras com mais de 108 CNVs, pois seriam consideradas amostras de baixa qualidade, portanto, seriam descartadas 75 amostras. Entretanto, optou-se por usar o valor de referência de 150 CNVs por amostra como limite máximo, porque assim seriam eliminadas menos amostras (53 amostras).

Como último parâmetro considerou-se o tamanho 5Mb como tamanho máximo para ser considerado como CNV. As CNVs com mais de 5Mb de comprimento foram excluídas, pois a confiança para inferência de CNVs a partir de matrizes de SNPs muito grandes é menor para comprimentos acima de 5Mb (comunicação verbal com LIU, 2017); não foi estabelecido

um tamanho mínimo para inclusão de CNVs porque acredita-se que quando é utilizado o programa CNVRuler uma CNV pequena pode fazer parte de uma CNVR (comunicação verbal com CHUD, 2017). O tamanho máximo obtido foi 7,5 Mpb e a variação de tamanho foi de 1 Kpb até 7,5 Mpb.

No momento de utilização do programa `filter_cnv`, após estabelecidos os critérios padrão (*default*) e os critérios adicionais, 359 amostras das 407 originais passaram no controle de qualidade, ou seja, 48 amostras foram removidas por não atenderem aos critérios adotados pelo programa. Com base nas estimativas calculadas, utilizou-se o pacote estatístico R no qual foram gerados, além do histograma com o NumCNV, também um histograma com a média de LRR (Gráfico 2), um gráfico de dispersão com a média de LRR por amostra (Gráfico 3) e um gráfico de dispersão com WF por amostra (Gráfico 4) para melhor compreensão dos dados. Os gráficos estão representados a seguir:

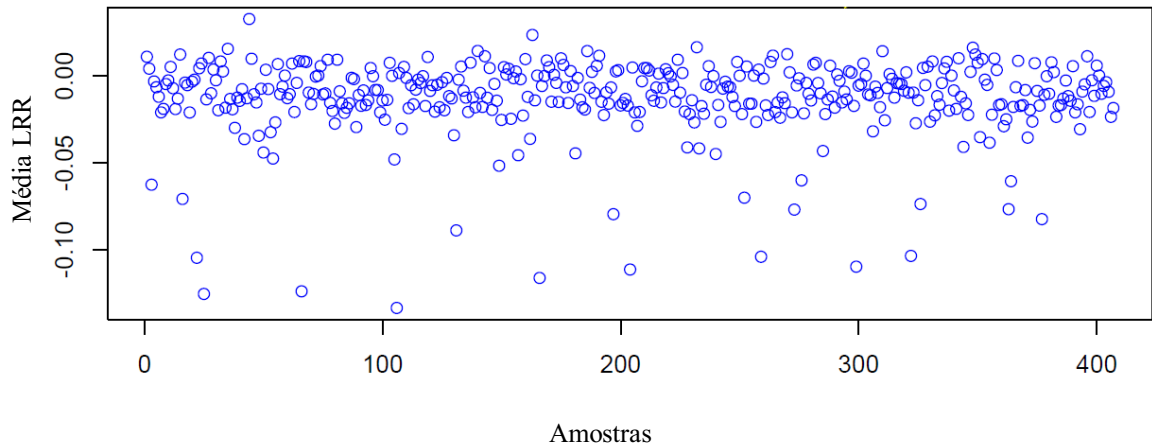
Gráfico 2. Histograma da média de LRR



Esse gráfico mostra a frequência da média do log de R calculada por amostra, portanto, espera-se grande frequência do neutro (mais amostras com média igual a zero), indicando a existência de duas cópias do SNP, o que é esperado em um organismo diplóide (2n).

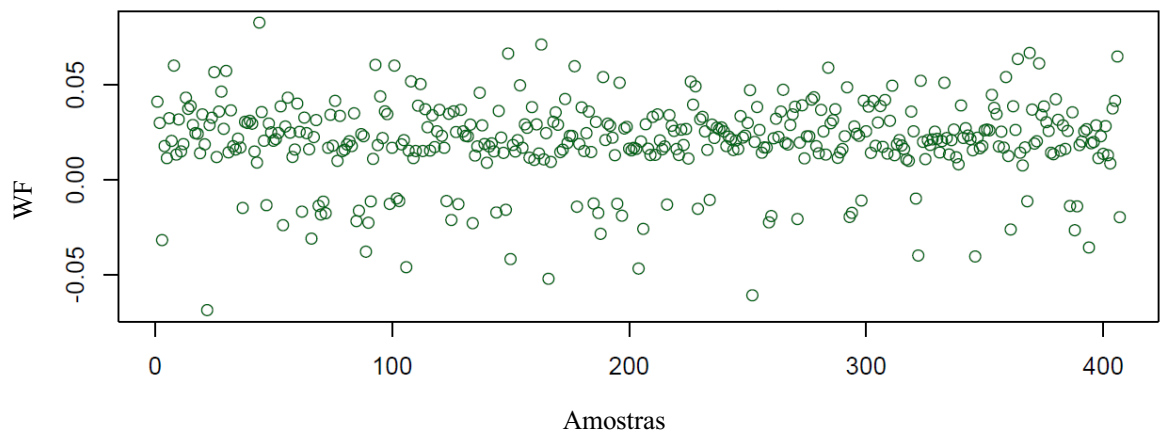
CNV representa deleção (perda) ou adição (ganho) de SNP em determinada posição. O máximo de inserções depende da capacidade de detecção do programa, no caso, o PennCNV consegue detectar até 4 cópias. Contudo, não é capaz de individualizar o sinal para dizer se ocorreu duplicação em ambos os cromossomos (AA AA) ou se em um deles existe triplicata e no outro uma cópia (AAA A).

Gráfico 3. Média de LRR por amostra



Quando a média de LRR está acima de zero significa que a amostra tem mais ganhos que perdas (média LRR alta), já quando a média de LRR está abaixo de zero a amostra tem mais perdas que ganhos (média LRR baixa).

Gráfico 4. WF por amostra



WF igual a zero é indicativo que não teve efeito do fator de onda. O gráfico mostra que nenhuma amostra obteve $WF = 0$, o que significa que o fator de ondulação teve efeito para todas as amostras. Quanto mais próximo de zero a amostra estiver, menor é o efeito do fator de onda, quanto mais distante maior o efeito, portanto, mais o ajuste será importante.